Jun 27th, 11:05 AM - 11:25 AM

Concurrent Sessions B: Case Studies - A Fisheries Information Management System (FIMS) for Petabyte Acoustic Telemetry

Brian LaMarche
Pacific Northwest National Laboratory

Christa Woodley
Pacific Northwest National Laboratory

Kenneth Auberry
Pacific Northwest National Laboratory

David Parrs
Pacific Northwest National Laboratory

Aaron Thronas
Pacific Northwest National Laboratory

See next page for additional authors

Presenter Information
Brian LaMarche, Christa Woodley, Kenneth Auberry, David Parrs, Aaron Thronas, Eric Choi, Brad Eppard, and Mark Weiland

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

A Fisheries Information Management System for Petabyte Scale Telemetry

BRIAN LAMARCHE, MARK WEILAND*, CHRISTA WOODLEY, ERIC CHOI, ADAM FLORY, NATHAN TRIMBLE, AARON PHILLIPS, DAVID PARRS, AARON THRONAS, MIKE RICHART, KEN AUBERRY
PACIFIC NORTHWEST NATIONAL LABORATORY

M. BRAD EPPARD
U.S. ARMY CORPS OF ENGINEERS, PORTLAND DISTRICT

Telemetry Studies at a Grand Scale

- Since 2004, the Juvenile Salmonid Acoustic Telemetry System (JSATS) has been used to study fish survival and passage on the Lower Columbia River

- 25,850 fish were tagged and released during 448 events between April-July (72 Days) in 2012.
  - Compared to 12,214 fish released in 2010 during 190 release events
  - Compared to 15,458 fish released in spring 2011 during 224 release events

- If the 2012 data was stored in its raw form, it would surmount a petabyte

- Large scale acoustic telemetry studies pose challenges for data
  - Management
  - Collection
  - Processing

- This work presents an information management approach for spatiotemporal analysis to support management of hydropower systems

Why? Survival Studies on the Lower Columbia

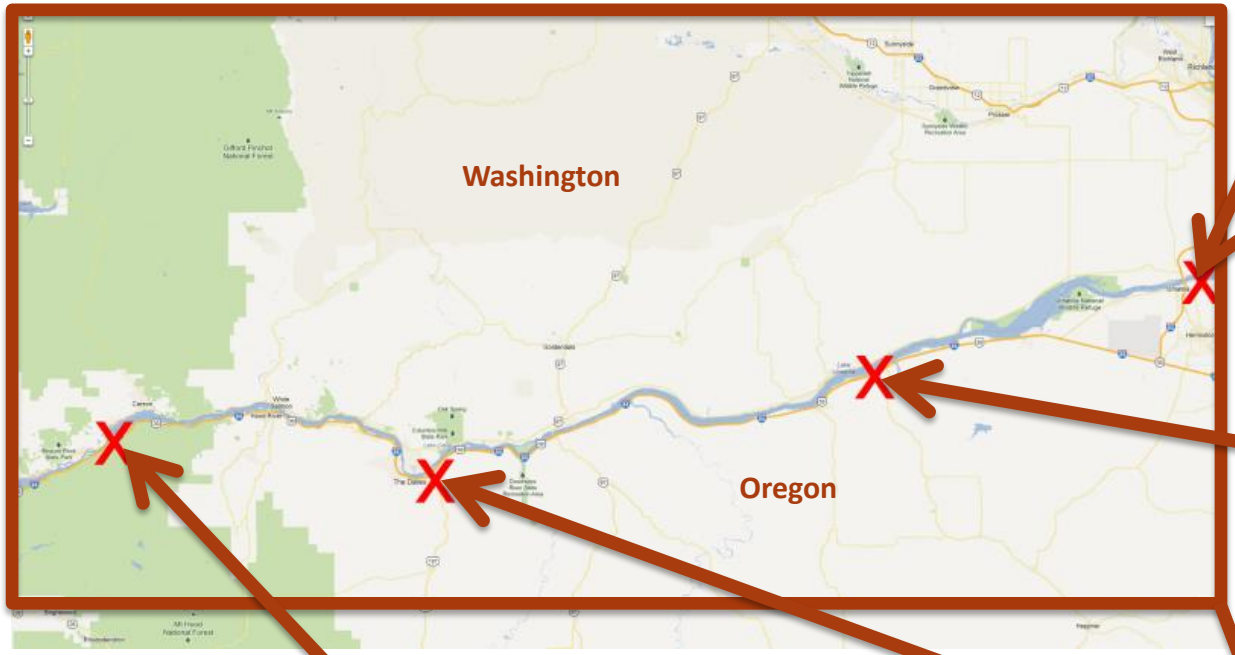
- Salmon are protected under the Endangered Species Act; Federal law requires the United States Army Corps of Engineers (USACE) evaluate survival of juvenile salmonids through the dams on the Columbia River*
  - Spring
    - 96% survival yearling Chinook salmon and steelhead
  - Summer
    - 93% survival sub yearling Chinook salmon
  - Criteria need to be met in 2 consecutive years with no slippage in metrics (forebay residence time, tailrace egress time, spill passage efficiency)

- Survival estimates and route of passage are estimated based on data collected using JSATS
- Route of passage helps the USACE:
  - understand how structural or hydraulic changes to dams can affect survival
  - determine the best routes to target for passage to increase survival

*Biological Opinion (NOAA BiOp 2008) and Columbia River Fish Accords

2012 Study Location – Lower Columbia River – 4 Dams

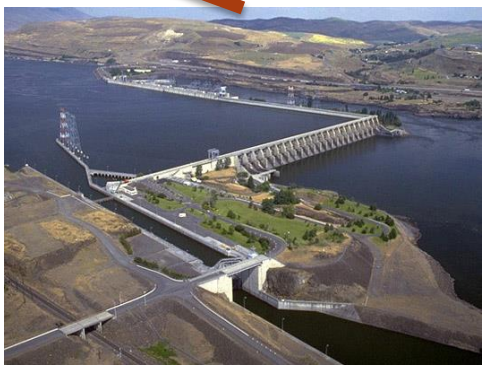
Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965



McNary Dam

John Day Dam

Bonneville Dam



The Dalles Dam

Scale and Challenges

- Study Area
  - 4 Dams (McNary, John Day, The Dalles, Bonneville Dams)
  - Study area extends over 417 rkm

- Telemetry Equipment
  - 84 autonomous nodes (14 arrays)
  - 4 dam mounted cabled arrays
    - 355 dam mounted hydrophones
    - 27 Miles of JSATS cable deployed
- Data Processing and Collection Resources
  - 92 acquisition computers at dams
  - 49+ computers for data analysis, status monitoring, and QA/QC
  - 33 Sierra Wireless Raven XE Airlink wireless modems and antenna for remote monitoring
- Over 30,000 fish handled, including 25,850 tagged and released
- Over 60 personnel working on 23 discrete activities
- QA/QC all data and procedures

What is a petabyte?

iPod 4G — 32 GB
X 32,000

iPod 4G — 32 GB
X 32

iPod 4G — 32 GB

1.44 MB

Text message

'h'

Byte

Kilobyte (KB)

Megabyte (MB)

Gigabyte (GB)

Terabyte (TB)

Petabyte (PB)

1e3

1e6

1e9

1e12

1e15

Fish

1 TB

Collection
500 MB
Tagging
10 GB
Condition
50 GB
Release
1.5 GB
Images
810 GB

Metadata

1 GB

Dam Operations
400 MB
Deployment
500 MB
Environmental
100 MB

Tag

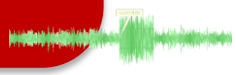
75 TB

Tag Activation / Code Space Management
0.5 GB

Tag Life
75,000 GB

~1,330,426 Gb

1.33 PB



Surgeon

50 GB

Surgeon Training

Surgeon Notes

Surgeon Feedback

QA/QC

5 TB

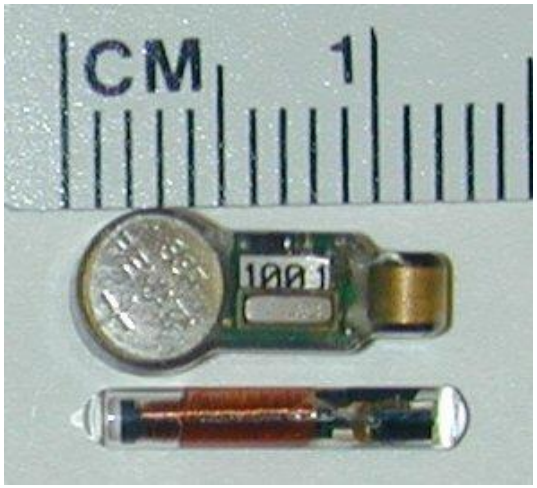
Error Analysis
5,000 GB
Other QA/QC Checks
25 GB

Detections

1.25 PB

Autonomous Array Data
250 GB
Cabled Array Data
1,250,000 GB

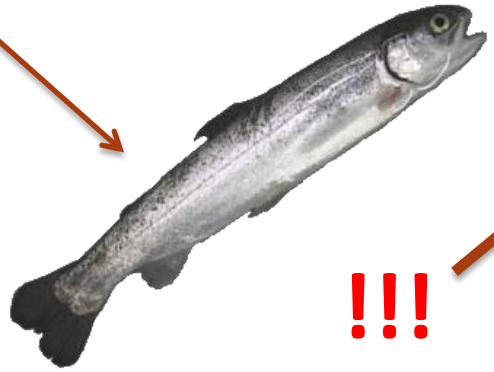
JSATS Acoustic Microtransmitter (AMT)



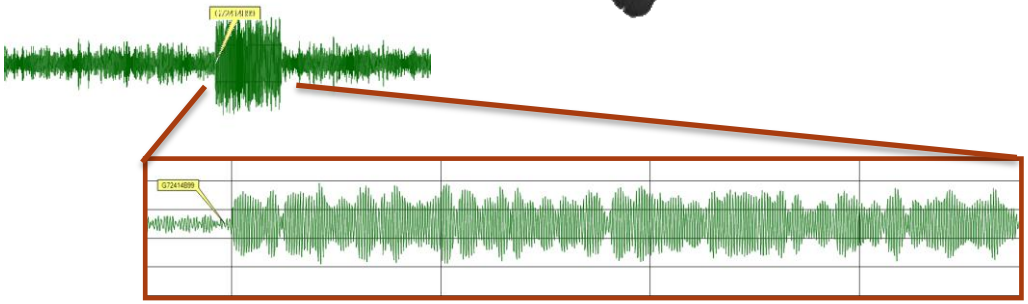
JSATS AMT (Active)

PIT (Passive)

- Signal
  - Acoustic
  - Binary Phase Shift Key (BPSK) encoded
  - 65,536 unique codes
  - Programmable ping rate (3s pri)
  - **167 kb file size per detection**



!!!



Acoustic Signal Emitted from tag

= 0111 0010 0010 0100…

Tagging

- Fish collected and tagged at the John Day Dam Smolt Monitoring Facility
- Fish quickly undergo several steps during tagging for documentation purposes prior to their release.
  - Photograph, fish condition, inventory
- Data acquired using suite of software tools called FishSuite!

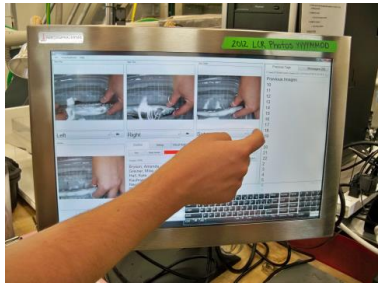
Tag → Photograph → Fish Condition → Inventory, recovery, and release

Tagging Process

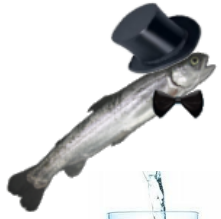
FishBooth! – Photography

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- At least 2 images are saved for EVERY Fish
  - 12 Megapixel resolution ~ 12 Mb (24 Mb per fish) ~
  - **810 GB – tagging only**
- Interfaces with digital SLR
- Saves information as EXIF meta-data in image file
  - PIT tag, Dam Location, Surgeon Name
- Integrates with PIT tag reader



Tag → **Photograph** → Fish Condition → Inventory, recovery, and release

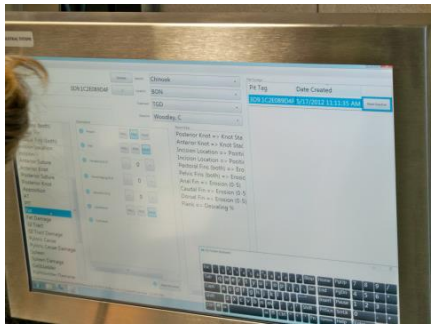


Tagging Process

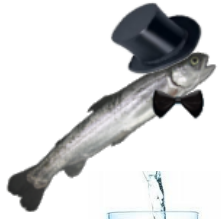
FishEye! – Assess fish condition rapidly

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- Quickly assesses fish condition (50+ metrics)
- Build customizable templates for anatomy / condition
- Saves data to local database
- Exports data for use with other tools / databases
- Integrates with scale and PIT tag reader



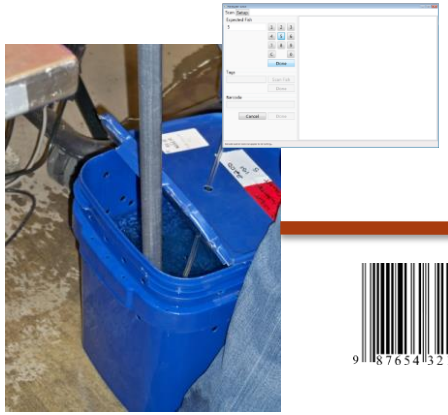
Tag → Photograph → **Fish Condition** → Inventory, recovery, and release



Tagging Process

FishBucket! and FreeWilly!

- Two applications that help track where fish are released and when

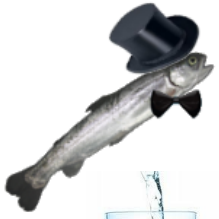


Inventory Fish Buckets using FishBucket (PIT Tag Reader)

Allow for recovery

Release, each bucket scanned with FreeWilly!, GPS locations stored

Tag → Photograph → Fish Condition → **Inventory, recovery, and release**



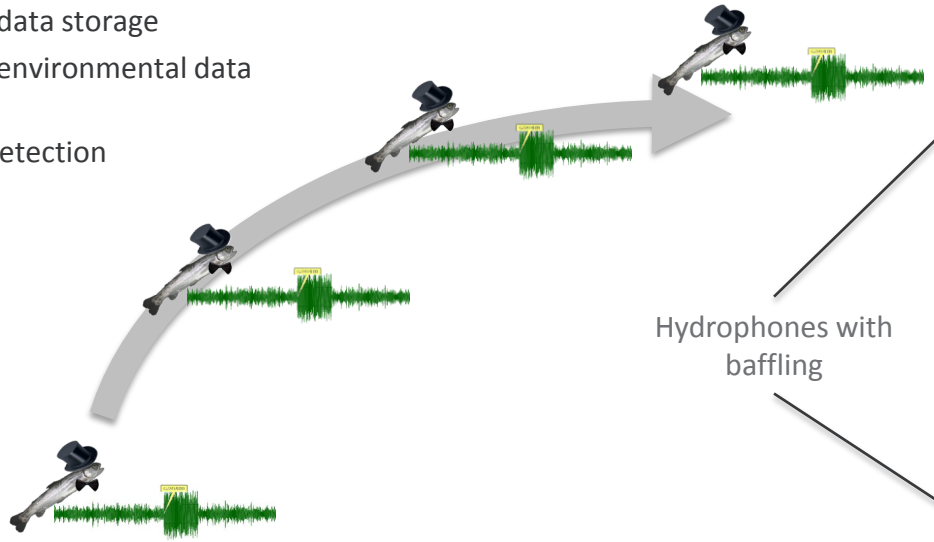
Tagging Process

Fish Detection and Acoustic Waveform Collection

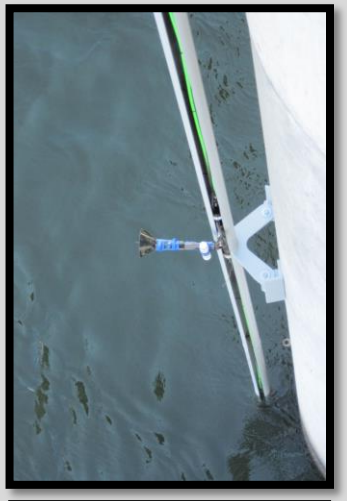
- JSATS Receivers – Cabled*
  - Hydrophones mounted on dam face (upstream)
  - Used to estimate 3D positions and tracks for route of passage

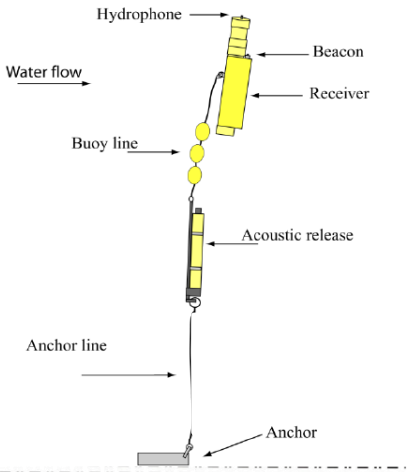
Pre-amplifier & Computers

- Autonomous Nodes
  - On-board power and data storage
  - Measures and stores environmental data (temp, depth, tilt)
  - Presences/Absence Detection



Hydrophones with baffling

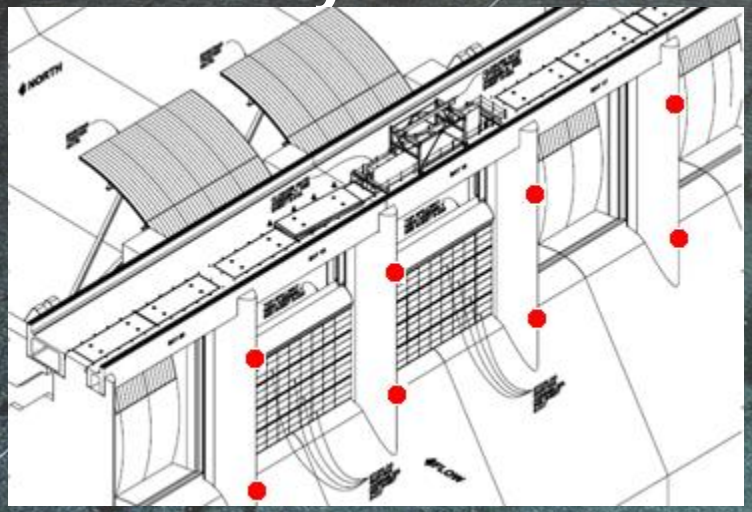


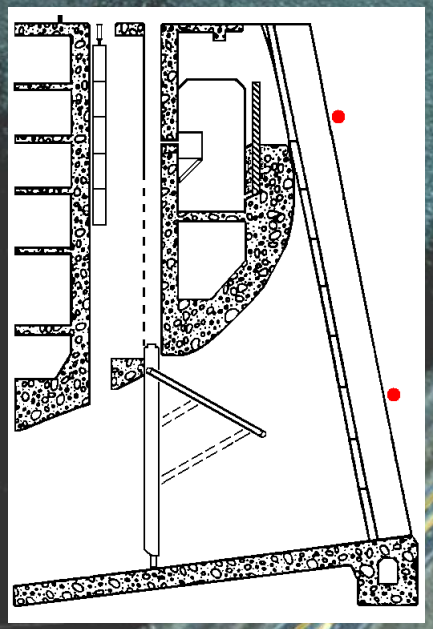


upriver

* Weiland, et. al., *Sensors 2011, 11, 5645-5660;*

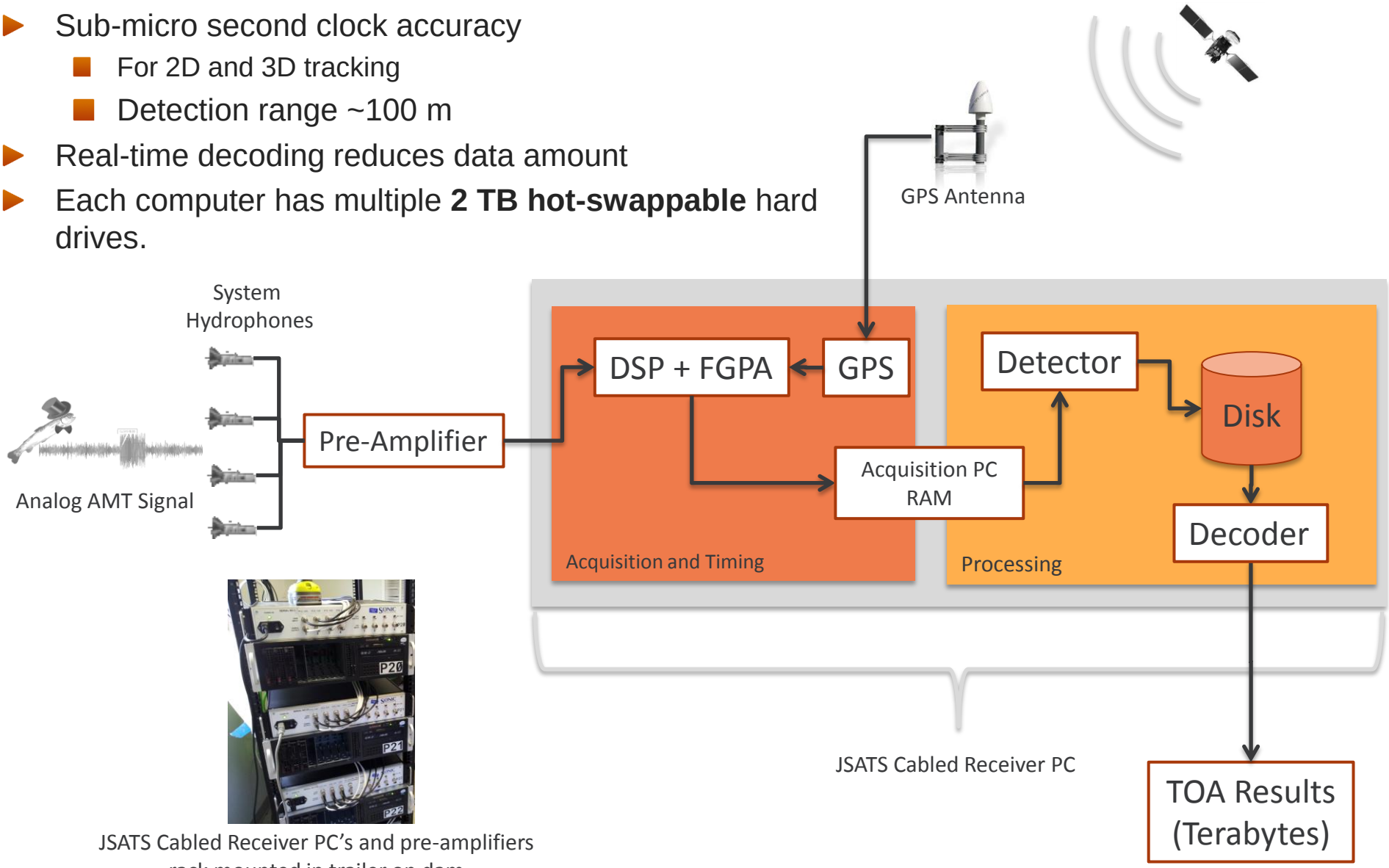
Hydrophone Locations – John Day Dam





JSATS Receivers - Cabled

- Sub-micro second clock accuracy
  - For 2D and 3D tracking
  - Detection range ~100 m
- Real-time decoding reduces data amount
- Each computer has multiple **2 TB hot-swappable** hard drives.

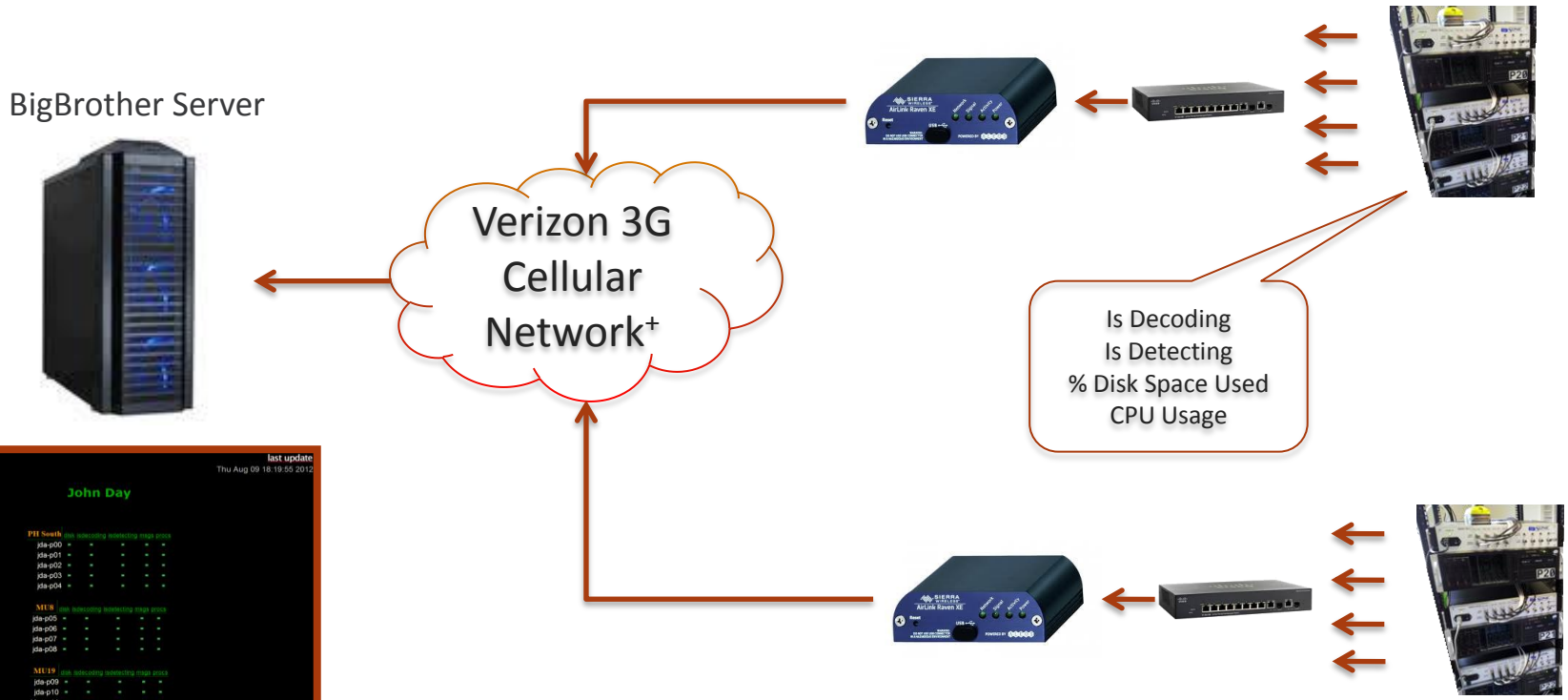


JSATS Cabled Receiver PC's and pre-amplifiers rack mounted in trailer on dam

Remote System Monitoring Mitigates Risk of System Downtime

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- System health is transmitted to a centralized server using wireless modems*+
- Health is displayed using BigBrother (Qwest Software)
- 375 Mb data usage per month / per system
- 33 deployed modems monitoring 92 systems



BigBrother web-page

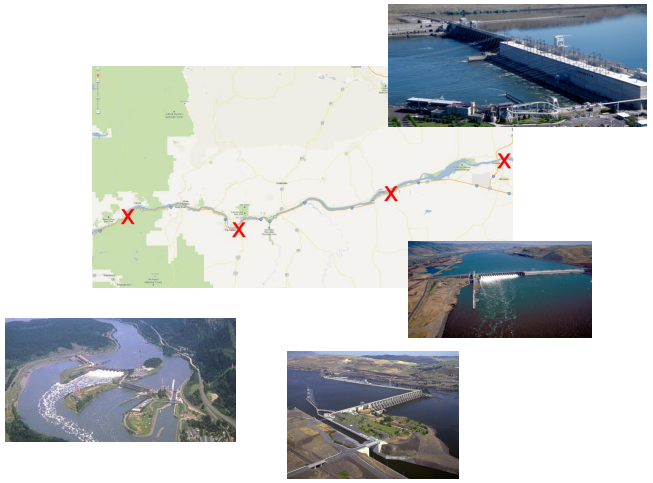
*Sierra Wireless Airlink Raven XE modem, +Satellite was used in remote locations.

Data Transfer and Unique Identification for Tracking

- **It is essential that data:**
  - Is given a unique identifier
  - formats are portable to allow data to be transferred to a centralized repository

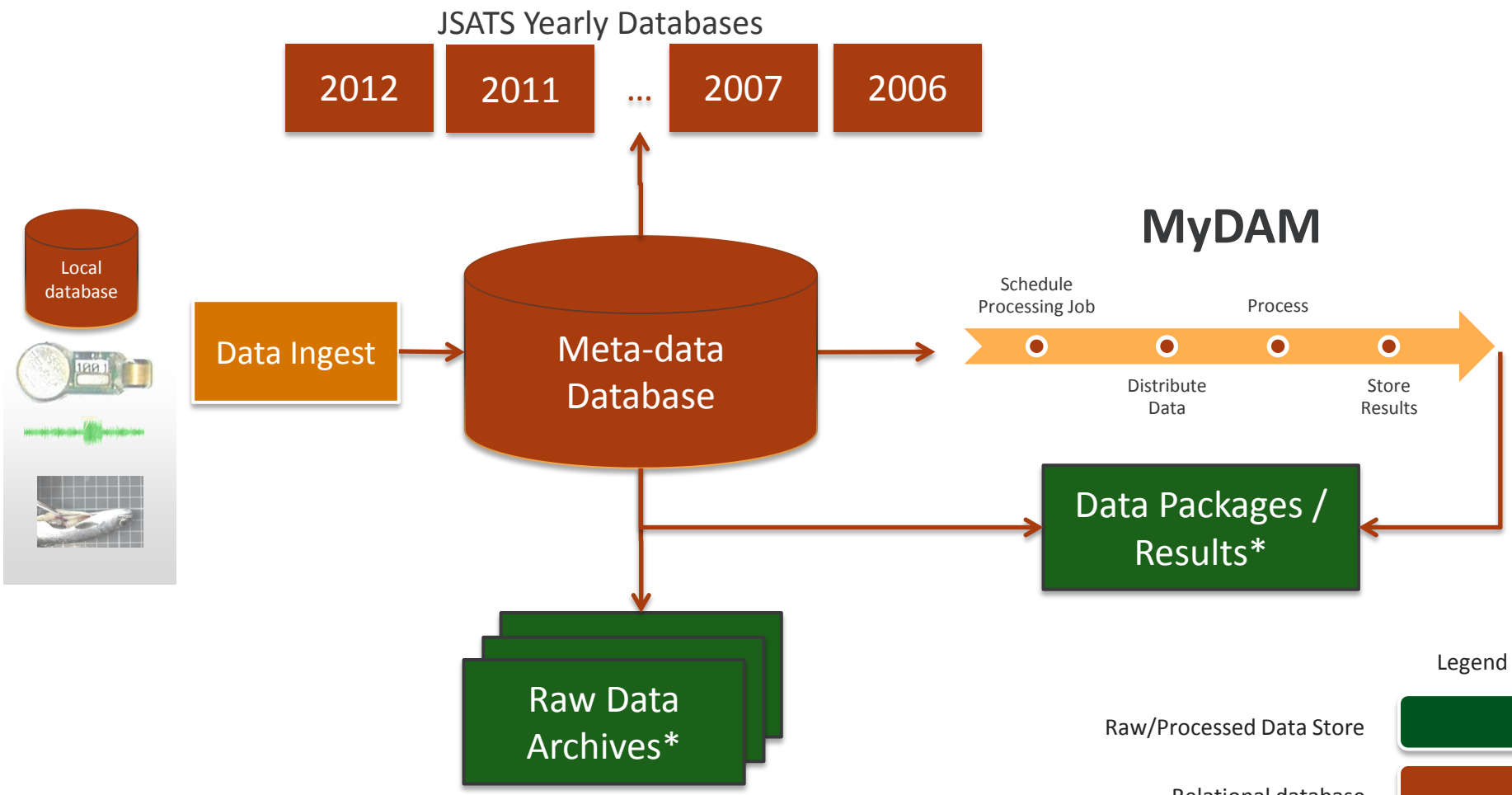
- **All tools were designed to write local databases or files**
  - SQLite
  - CompactSQL
  - Comma separated value (CSV)

- A few Unique ID options
  - Globally unique identifier — GUID
    - 128-bit value, with 2^{122} possibilities
    - Images, surgery notes, observations
  - Time / Date / Location
    - Detections
    - Releases



Fisheries Information Management System Architecture

- Based on management information system from proteomics+

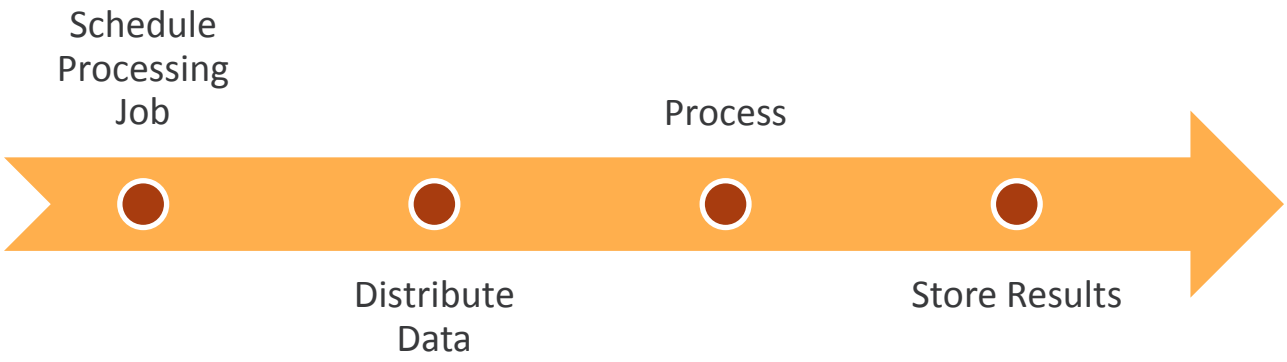


*Raided for data protection

+ Kiebel, et. al., "PRISM: a data management system for high-throughput proteomics." Proteomics, 2006. 6(6). 1783-1790.

Data Processing Pipeline

- Automates data processing and tracks how a tool is run
- MyDAM
  - Route of passage
  - Gap check analysis
  - JSATS Detection
- Investigating Hadoop as processing pipeline manager
  - Distributed / Parallel processing using MapReduce
  - Used by Twitter, Facebook, Yahoo!, NASA
  - Manages job tracking / restarts
- **Move away from saving raw waveforms!!**



Future and Conclusion

- Telemetry at the petabyte scale introduces data management, collection, and processing challenges
  - Remote system monitoring is useful for maintaining full operation
- Building data processing pipelines in addition to database management systems is useful for reproducibility and automated analysis
- Introduce new processing pipelines and external data sources
- Add a web interface to facilitate data analysis and interaction with FIMS
- Perform meta-data analysis relating dam operations to route of passage and survival
  - Answer questions posed by hydropower management

Acknowledgements

- USACE Portland District
- PNNL Staff
  - Tom Carlson
  - Jina Kim
  - Daniel Deng
  - Gene Ploskey
  - James Hughes
  - Eric Fischer
  - Shon Zimmerman
  - Katie Wagner
  - Charles Pratt
  - Jeni Smith (Sequim)
  - Valorie Hiner
  - Tylor Abel
  - Yuan Yong
- Pacific States
  - Scott Carpenter
  - George Batten
  - Aaron Cushing